

# **Defendant's Exhibit 25**

GIBSON, DUNN & CRUTCHER LLP  
Orin Snyder (*pro hac vice*)  
osnyder@gibsondunn.com  
200 Park Avenue  
New York, NY 10166-0193  
Telephone: 212.351.4000  
Facsimile: 212.351.4035

GIBSON, DUNN & CRUTCHER LLP  
Deborah Stein (SBN 224570)  
dstein@gibsondunn.com  
333 South Grand Avenue  
Los Angeles, CA 90071-3197  
Telephone: 213.229.7000  
Facsimile: 213.229.7520

Kristin A. Linsley (SBN 154148)  
klinsley@gibsondunn.com  
Martie Kutscher (SBN 302650)  
mkutscherclark@gibsondunn.com  
555 Mission Street, Suite 3000  
San Francisco, CA 94105-0921  
Telephone: 415.393.8200  
Facsimile: 415.393.8306

Joshua S. Lipshutz (SBN 242557)  
jlipshutz@gibsondunn.com  
1050 Connecticut Avenue, N.W.  
Washington, DC 20036-5306  
Telephone: 202.955.8500  
Facsimile: 202.467.0539

*Attorneys for Defendant Facebook, Inc.*

UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION

IN RE: FACEBOOK, INC. CONSUMER  
PRIVACY USER PROFILE LITIGATION,

This document relates to:

ALL ACTIONS

CASE NO. 3:18-MD-02843-VC

**MOTION FOR RECONSIDERATION OF  
SPECIAL MASTER'S ORDER  
REGARDING NAMED PLAINTIFF DATA**

## I. Introduction

Facebook respectfully seeks reconsideration of the Special Master’s November 29, 2021 Order re: Plaintiffs’ Motion to Compel Production of Plaintiff Data (the “tentative order”). The tentative order, if it became final, would require Facebook to conduct an onerous investigation of all of its data sources that, if even possible, would take years to complete. When Facebook initially raised this infeasibility to Judge Corley, Plaintiffs *insisted* that they were not seeking user data relating to Named Plaintiffs *unless it was shared or made accessible* to third parties, and they assured the Court that this limitation would *narrow the information Facebook had to produce*. The focus then turned to whether Facebook had produced the three categories of user data that Judge Corley found potentially relevant, with the understanding that it could be relevant only if it met the “shared or made accessible” test that Plaintiffs themselves advanced.

The Special Master’s tentative order threatens to undo this forward progress and send the parties back to square one. Facebook thus seeks reconsideration of the tentative order (1) to the extent that it unwinds more than one year of negotiations, litigation, and the parties’ agreement by expanding relevant Named Plaintiff data *beyond* data that was shared or made accessible, and (2) because compliance with the tentative order as written, particularly within the specified timeframe, is not feasible, as set forth in the accompanying declarations of Mengge Ji and David Pope.

## II. Facebook seeks reconsideration of the Special Master’s finding that data relating to the Named Plaintiffs is discoverable even if not shared.

Facebook seeks reconsideration of the Special Master’s finding that “Discovery Order No. 9 does not limit the scope of discoverable data related to the Named Plaintiffs to data that was shared with third parties . . . because Judge Corley’s ruling contains no language indicating such a limitation.” Order ¶ 15 (Exhibit G).

1. Discovery Order No. 9 *does* state that the user data relevant to this case is limited to shared data. But Discovery Order No. 9 largely addressed a separate issue. In the briefs underlying Discovery Order No. 9, Facebook argued that relevant user data is data that users posted *on Facebook*. Judge Corley rejected this position on the basis that relevant data includes all data that Facebook **shares** with third parties (whether posted on Facebook or not). Judge Corley wrote: “Plaintiffs correctly argue

1 that Facebook’s restrictive view of relevant discovery would exclude an enormous amount of  
 2 information that Facebook collects and **shares** with third parties about Facebook’s users. The district  
 3 court’s order (Dkt. No. 298) did not limit Plaintiffs’ claims to only challenging the **sharing** of data  
 4 Facebook collects from a user’s on-platform activity; the claims also challenge Facebook’s **sharing** of  
 5 user data and alleged failure to monitor how third parties used such **shared** information.” Dkt. 557 at  
 6 1–2 (emphases added) (Exhibit I). Judge Corley went on to describe three categories of discoverable  
 7 data that included both on-Facebook and off-Facebook activities. *Id.* at 2. Her description of these  
 8 categories of data did not reverse her prior language clarifying that the scope of Plaintiffs’ live claims—  
 9 and therefore the scope of relevant discovery—concerns “Facebook’s sharing of user data.” *Id.*

10 2. Discovery Order No. 9 did not dedicate significant space to clarifying that the scope of  
 11 discoverable user data is data that Facebook shared with third parties because the parties agreed on  
 12 this point. In the briefing underlying Discovery Order No. 9, Plaintiffs stated: “Plaintiffs seek only a  
 13 holding that the sensitive data Facebook collected about *ten Named Plaintiffs* and *shared* with third  
 14 parties is relevant.” Dkt. 547-3 at 9 (emphasis in original) (Exhibit J).<sup>1</sup> Plaintiffs explicitly represented  
 15 that their narrowed request would narrow the scope of discovery: “Plaintiffs do not contend that  
 16 information that was not shared is relevant, which substantially narrows the information Facebook  
 17 would be required to produce in this case.” *Id.* (emphasis added). In particular, Plaintiffs told Judge  
 18 Corley that they did “not demand . . . ‘that Facebook search *millions* of disaggregated data sets for any  
 19 data to have ever crossed Facebook’s systems relating to a Named Plaintiff and any derivative materials  
 20 drawing on that data.’” *Id.* (quoting Facebook’s opposition brief). The Special Master’s tentative  
 21 order, however, would require exactly that.

22 3. At the December 9, 2020 hearing before Judge Corley on this subject, counsel for  
 23 Plaintiffs reiterated that the information that they sought about Named Plaintiffs was “information

24 <sup>1</sup> Plaintiffs repeated this position numerous times in their briefing. *See* Dkt. 547-3 at 1 (“This discovery  
 25 dispute concerns sensitive user information that Facebook has shared with third parties without users’  
 26 consent.”); *id.* at 2 (“[S]ensitive user information is relevant if Facebook shared it without users’  
 27 consent.”); *id.* at 4 (“[T]he legal theories upheld at the pleading stage” turn on “whether Facebook  
 28 shared [sensitive information] with third parties.”); *id.* at 5 (“Plaintiffs have standing . . . because their  
 sensitive information was disseminated to third parties in violation of their privacy.” (quotation marks  
 and citation omitted)); *id.* at 9 (“Plaintiffs seek an order holding that all sensitive data about the ten  
 Named Plaintiffs that Facebook shared with or made accessible to third parties is relevant to this  
 action.”).

shared or made accessible.” Opp. Ex. R at 18:15–16 (Exhibit F) (argument by Counsel Loeser to Judge Corley).

4. In briefing before the Special Master, Plaintiffs again conceded: “Plaintiffs have always sought, and continue to seek, content and information that has been shared with or made accessible to third parties. And they are not arguing that content and information is relevant if Facebook did not share it with or make it accessible to third parties.” Plaintiffs’ Reply at 1 (Exhibit H).

5. By declining to “limit the scope of discoverable data related to the Named Plaintiffs to data that was shared with third parties,” Order ¶ 15, the Special Master’s tentative order could inject the extremely burdensome discovery that the parties and Judge Corley have already recognized is not at issue. To the extent the Special Master meant only that discoverable information included information “shared *or made accessible*”—which would be consistent with Plaintiffs’ statements and Judge Corley’s orders—Facebook respectfully requests that the tentative order be clarified to say so.

6. The Special Master’s order also cites Discovery Order No. 11 as evidence that “Judge Corley clarified that Facebook’s interpretation of Discovery Order No. 9 is not what Judge Corley intended.” Order ¶ 16. Judge Corley did not issue Discovery Order No. 11 in order to clarify or expand her prior order. Judge Corley issued Discovery Order No. 11 to describe the scope of a 30(b)(6) deposition that she authorized at a December 9, 2020 discovery hearing. Dkt. 588 at 1 (Exhibit K). The purpose of that deposition was to allow Plaintiffs to explore whether any data Facebook shares with third parties had not been produced for the Named Plaintiffs. Facebook had already explained that it does not share inferences about users with third parties. *See* Opp. Ex. R. at 20:15–18.<sup>2</sup> But Plaintiffs expressed “disbelief” in Facebook’s representations about what data was shared with third parties, and demanded a deposition to verify those representations: “We just don’t believe . . . their description of **what is or is not shared or made accessible**. We need to put somebody under oath and have them testify about that.” *Id.* at 26:7–9, 28:15–17 (emphasis added). Judge Corley therefore

<sup>2</sup> This is for a basic technological reason. When a third party—such as an app developer or business partner—obtains user-related data, it accesses it through an “application programming interface,” or “API.” Opp. Ex. E ¶¶ 3–7 (Exhibit E). The APIs at issue pulled information from Facebook’s “Social Graph,” not data warehouses like Hive. *See id.* A user’s DYI file, which Facebook produced for all Named Plaintiffs, contains the most complete current set of data about that user that is in the Social Graph (and more). *See* Opp. Ex. C ¶ 5 (Exhibit D).

1 allowed Plaintiffs to conduct a 30(b)(6) deposition for that purpose: “to verify the representation that  
 2 yes, we collect this information—inferential data, but it is not made accessible to third parties.” *Id.* at  
 3 35:3–5.

4 7. Discovery Order No. 11 describes the scope of the deposition Judge Corley authorized;  
 5 it does not alter Discovery Order No. 9. The statement from Discovery Order No. 11 that the Special  
 6 Master quoted in his order confirms this: “How [Facebook] specifically uses this data is an open  
 7 question, but if the Court were to accept Facebook’s arguments about the scope of production, it would  
 8 eliminate Discovery Order No. 9’s third category of discovery: data inferred from a user’s on or off-  
 9 platform activity.” Order ¶ 16 (quoting Discovery Order No. 11). That statement confirms that *if*  
 10 *Facebook did not share inferred data*—the question that the 30(b)(6) deposition was meant to verify—  
 11 *then the third category of data would not be discoverable*. Respectfully, the Special Master’s reading  
 12 of this statement to mean that non-shared data is discoverable is incorrect and contrary to Discovery  
 13 Order No. 9, the briefing underlying it, Discovery Order No. 11, and the entire purpose of the 30(b)(6)  
 14 deposition that Judge Corley authorized.

15 **III. In any event, Facebook seeks reconsideration of the Special Master’s order regarding**  
 16 **what information Facebook must produce.**

17 Even if the Special Master does not reconsider his finding that “Discovery Order No. 9 does  
 18 not limit the scope of discoverable data related to the Named Plaintiffs to data that was shared with  
 19 third parties,” Order ¶ 15, Facebook respectfully requests reconsideration of the portion of the Special  
 20 Master’s tentative order regarding the information Facebook must produce at this time. Facebook is  
 21 prepared to “provide a list of data sources that may,” to the best of Facebook’s current knowledge,  
 22 “contain information related to” Facebook users, which may include “the Named Plaintiffs.” Order  
 23 ¶ 19. But complying with the other aspects of the Special Master’s order is not feasible, particularly  
 24 within the proposed timeframe. Facebook has good cause to raise each of these issues in a motion for  
 25 reconsideration because (i) Facebook understood that the parties and Court were in agreement that  
 26 information that was not “shared or made accessible” to third parties was not discoverable, *see* Ex. A  
 27 (Kutcher Decl.) ¶¶ 1-6, and (ii) the Special Master’s Order requires Facebook to provide information  
 28 Plaintiffs did not request in their Motion to Compel, *id.* at ¶ 7.

8. *First*, Facebook requests reconsideration of the Special Master’s order that it provide “the name of [all] database[s] or data log[s]” that “may contain information related to the Named Plaintiffs,” Order ¶ 19, as Facebook does not presently have a full list of these sources and is unable to prepare one within a reasonable time period. Facebook has already spent more than one year creating an inventory of its data systems, an effort that required coordination across dozens of different Facebook teams and hundreds of Facebook employees and is not yet complete. *See* Ex. B ¶ 3 (Pope Decl.). Facebook is also working to identify the specific data assets within those systems, including the data logs the Special Master ordered to be identified, but Facebook must develop new tools to inventory and understand this information. *See id.* ¶ 10. Facebook is providing with this motion a list of the data systems it has identified to date as storing or interacting with user data (without accounting for whether that data is aggregated or de-identified from a user’s account), but this project is not yet complete and Facebook cannot yet identify all of the “data logs” within each of the data systems identified. *See id.* ¶¶ 3–4, 6; *id.* Ex. A.

9. *Second*, Facebook requests reconsideration of the Special Master’s order that it provide “a description of [each] data source’s purpose and function.” Order ¶ 19. Facebook has not compiled descriptions of every data system’s “purpose and function,” which may vary depending on which Facebook team is using it (and many data systems are used by multiple teams). *See* Ex. B ¶ 7 (Pope Decl.). Compiling this information about all of Facebook’s data systems would require Facebook to repeat the same process it used to identify the data systems in the first place, including more than one year of revisiting each data system with the relevant Facebook teams and employees. *See id.* ¶ 9.

10. *Third*, for the same reasons, Facebook requests reconsideration of the Special Master’s order that it provide “a description of the types of Named Plaintiff data contained in the data source.” Order ¶ 19. Facebook has not compiled this information, which can vary depending on the team using the data system, and compiling this information would again require more than one year of revisiting each data system with the relevant Facebook teams and employees. *See* Ex. B ¶¶ 8–9 (Pope Decl.).

11. *Fourth*, Facebook requests reconsideration of the Special Master’s order that the parties submit “a proposed protocol for the production of Named Plaintiffs’ data from the data sources identified by Facebook.” Order ¶ 20. For multiple reasons, broad production of all additional data

1 related to the Named Plaintiffs is extremely burdensome, if not impossible, as Facebook's systems are  
2 not generally designed to allow for the identification and export of data related to a specific user. *See*  
3 Ex. C ¶¶ 10–31 (Ji Decl.). Blanket production of all data from just one additional data system, Hive,  
4 would require a manual search of millions of data tables over the course of hundreds of years of full-  
5 time work. *See id.* And Facebook has determined that its existing tools cannot be leveraged to identify  
6 specific data for retrieval; rather, Facebook would need to dedicate a full team of engineers for more  
7 than a year to develop retrieval tools. *See* Ex. B ¶¶ 10–11 (Pope Decl.). Moreover, any data ultimately  
8 retrieved as a result of these massive efforts would likely be irrelevant and unusable, as Facebook's  
9 raw data is unintelligible outside of Facebook. *See* Ex. C ¶¶ 10, 23 (Ji Decl.).

#### 10 **IV. Conclusion**

11 Facebook respectfully asks the Special Master to reconsider his finding that data relating to the  
12 Named Plaintiffs is discoverable even if not shared. Even if the Special Master does not reconsider  
13 that finding, Facebook respectfully requests that the Special Master reconsider what information  
14 Facebook must produce.



1 Dated: December 10, 2021

**GIBSON, DUNN & CRUTCHER, LLP**

2 By: /s/ Deborah Stein

3 Orin Snyder (*pro hac vice*)

4 osnyder@gibsondunn.com

200 Park Avenue

5 New York, NY 10166-0193

Telephone: 212.351.4000

Facsimile: 212.351.4035

6 Deborah Stein (SBN 224570)

7 dstein@gibsondunn.com

333 South Grand Avenue

8 Los Angeles, CA 90071-3197

Telephone: 213.229.7000

9 Facsimile: 213.229.7520

10 Joshua S. Lipshutz (SBN 242557)

11 jlipshutz@gibsondunn.com

1050 Connecticut Avenue, N.W.

12 Washington, DC 20036-5306

Telephone: 202.955.8500

Facsimile: 202.467.0539

13 Kristin A. Linsley (SBN 154148)

14 klinsley@gibsondunn.com

15 Martie Kutscher (SBN 302650)

mkutscherclark@gibsondunn.com

555 Mission Street, Suite 3000

16 San Francisco, CA 94105-0921

Telephone: 415.393.8200

17 Facsimile: 415.393.8306

18 *Attorneys for Defendant Facebook, Inc.*

# **EXHIBIT A**

GIBSON, DUNN & CRUTCHER LLP  
Orin Snyder (*pro hac vice*)  
osnyder@gibsondunn.com  
200 Park Avenue  
New York, NY 10166-0193  
Telephone: 212.351.4000  
Facsimile: 212.351.4035

GIBSON, DUNN & CRUTCHER LLP  
Deborah Stein (SBN 224570)  
dstein@gibsondunn.com  
333 South Grand Avenue  
Los Angeles, CA 90071-3197  
Telephone: 213.229.7000  
Facsimile: 213.229.7520

Kristin A. Linsley (SBN 154148)  
klinsley@gibsondunn.com  
Martie Kutscher (SBN 302650)  
mkutscherclark@gibsondunn.com  
555 Mission Street, Suite 3000  
San Francisco, CA 94105-0921  
Telephone: 415.393.8200  
Facsimile: 415.393.8306

Joshua S. Lipshutz (SBN 242557)  
jlipshutz@gibsondunn.com  
1050 Connecticut Avenue, N.W.  
Washington, DC 20036-5306  
Telephone: 202.955.8500  
Facsimile: 202.467.0539

*Attorneys for Defendant Facebook, Inc.*

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION**

IN RE: FACEBOOK, INC. CONSUMER  
PRIVACY USER PROFILE LITIGATION,

This document relates to:  
  
ALL ACTIONS

CASE NO. 3:18-MD-02843-VC

**DECLARATION OF MARTIE  
KUTSCHER IN SUPPORT OF  
FACEBOOK, INC.'S MOTION FOR  
RECONSIDERATION OF SPECIAL  
MASTER'S ORDER REGARDING  
NAMED PLAINTIFF DATA**

1 I, Martie Kutscher, hereby declare as follows:

2 1. I am an associate at the law firm of Gibson, Dunn & Crutcher LLP, counsel of record  
3 for Facebook, Inc. (“Facebook”) in the above-captioned matter. I am a member in good standing of  
4 the State Bars of California, New Jersey, and New York. I submit this declaration in support of  
5 Facebook’s Motion for Reconsideration of the Special Master’s Order Regarding Named Plaintiff Data.  
6 I make this declaration on my own knowledge, and I would testify to the matters stated herein under  
7 oath if called upon to do so.

8 2. In briefing before Judge Corley in October 2020, Plaintiffs wrote: “Plaintiffs seek only  
9 a holding that the sensitive data Facebook collected about *ten Named Plaintiffs* and *shared* with third  
10 parties is relevant. Plaintiffs do not contend that information that was not shared is relevant, which  
11 substantially narrows the information Facebook would be required to produce in this case.” Dkt. 547-  
12 3 at 9 (emphasis in original). Plaintiffs also stated that they did “not demand . . . ‘that Facebook search  
13 *millions* of disaggregated data sets for any data to have ever crossed Facebook’s systems relating to a  
14 Named Plaintiff and any derivative materials drawing on that data—such as data sets tracking hours of  
15 peak user activity to monitor strains on Facebook’s system.’” *Id.* (quoting Facebook’s opposition  
16 brief).

17 3. Plaintiffs also stated at a hearing before Judge Corley that the information they sought  
18 about Named Plaintiffs was “information shared or made accessible” to third parties. Opp. Ex. Q at  
19 18:15–16.

20 4. Judge Corley’s order similarly reflected the parties’ agreement that discoverable  
21 information was limited to information that was shared or made accessible to third parties. *See* Opp.  
22 Ex. P (“The district court’s order (Dkt. No. 298) did not limit Plaintiffs’ claims to only challenging the  
23 **sharing** of data Facebook collects from a user’s on-platform activity; the claims also challenge  
24 Facebook’s **sharing** of user data and alleged failure to monitor how third parties used such **shared**  
25 information.”) (Discovery Order No. 9) (emphases added); *see also, e.g.*, Opp. Ex. R at 35:3–5 (stating  
26 that the purpose of the Rule 30(b)(6) deposition she authorized was “to verify the representation that  
27 yes, we collect this information—inferential data, but it is not made accessible to third parties”) (Dec.  
28 9, 2020 hearing).

1           5.       Based on these and other similar statements from Plaintiffs and Judge Corley, Facebook  
2 understood that the parties and Court were in agreement that information that was not “shared or made  
3 accessible” to third parties was not discoverable.

4           6.       Because Facebook understood the parties to agree on this point, Facebook’s initial  
5 briefing before the Special Master did not detail the extremely burdensome nature or impossibility of  
6 identifying and producing all data related to the Named Plaintiffs. In light of the Special Master’s  
7 tentative order, Facebook provides greater detail on those issues in its motion for reconsideration and  
8 accompanying declarations.

9           7.       The Special Master’s tentative order also requires Facebook to provide “a description  
10 of the types of Named Plaintiff data contained in [each] data source.” Order ¶ 19. Plaintiffs did not  
11 request such a description in their motion to compel, so Facebook did not discuss the burden of  
12 providing such a description in its initial briefing before the Special Master. In light of the Special  
13 Master’s tentative order, Facebook provides greater detail on this issue in its motion for reconsideration  
14 and accompanying declarations.

15           8.       I declare under penalty of perjury under the laws of the United States of America that  
16 the foregoing is true and correct.

17  
18 Executed on December 10, 2021 in Palo Alto, California.

19  
20 

21 Martie Kutscher  
22  
23  
24  
25  
26  
27  
28

## **EXHIBIT B**

GIBSON, DUNN & CRUTCHER LLP  
Orin Snyder (*pro hac vice*)  
osnyder@gibsondunn.com  
200 Park Avenue  
New York, NY 10166-0193  
Telephone: 212.351.4000  
Facsimile: 212.351.4035

Kristin A. Linsley (SBN 154148)  
klinsley@gibsondunn.com  
Martie Kutscher (SBN 302650)  
mkutscherclark@gibsondunn.com  
555 Mission Street, Suite 3000  
San Francisco, CA 94105-0921  
Telephone: 415.393.8200  
Facsimile: 415.393.8306

GIBSON, DUNN & CRUTCHER LLP  
Deborah Stein (SBN 224570)  
dstein@gibsondunn.com  
333 South Grand Avenue  
Los Angeles, CA 90071-3197  
Telephone: 213.229.7000  
Facsimile: 213.229.7520

Joshua S. Lipshutz (SBN 242557)  
jlipshutz@gibsondunn.com  
1050 Connecticut Avenue, N.W.  
Washington, DC 20036-5306  
Telephone: 202.955.8500  
Facsimile: 202.467.0539

*Attorneys for Defendant Facebook, Inc.*

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION**

IN RE: FACEBOOK, INC. CONSUMER  
PRIVACY USER PROFILE LITIGATION,

This document relates to:

ALL ACTIONS

CASE NO. 3:18-MD-02843-VC

**DECLARATION OF DAVID POPE IN  
SUPPORT OF FACEBOOK, INC.'S  
MOTION FOR RECONSIDERATION OF  
THE SPECIAL MASTER'S ORDER RE  
PLAINTIFFS' MOTION TO COMPEL  
PRODUCTION OF PLAINTIFF DATA**

Discovery Special Master Daniel Garrie, Esq.

**DECLARATION OF DAVID POPE IN SUPPORT OF FACEBOOK, INC.'S  
MOTION FOR RECONSIDERATION OF THE SPECIAL MASTER'S ORDER RE  
PLAINTIFFS' MOTION TO COMPEL PRODUCTION OF PLAINTIFF DATA**

I, David Pope, declare:

1. I am a Group Technical Program Manager on the Core Infra Team at Meta, Inc. f/k/a Facebook, Inc. (hereinafter “Facebook”). I make this declaration on my own knowledge, and I would testify to the matters stated herein under oath if called upon to do so.

2. In my role as a Group Technical Program Manager, I am responsible for supporting the Core Infra Engineering teams. Through my work at Facebook, I am familiar with Facebook’s efforts to inventory its data systems and the data assets (*i.e.*, individual logs, data sets, or other units of data) within them.

3. My team and others have been working to inventory all of the data systems (*i.e.*, data storage or analysis tools) within Facebook and understand which of them retain user data. In order to do so, we have spent more than a year working with dozens of different Facebook teams and hundreds of Facebook employees to identify the data systems used by the company. This work is still ongoing.

4. A list of the data systems we have identified as storing or interacting with user data is attached hereto as **Exhibit A**.

5. I understand that the Special Master in this matter has requested that Facebook produce “a list of data sources that may contain” user data and for each source provide (1) “the name of the data source or data log”; (2) “a description of the data source’s purpose and function”; and (3) “a description of the types of [user data] . . . contained in the data source.”

6. The inventory my team has compiled does not include all “data logs” within each of the data systems identified.



7. My team also has not compiled a comprehensive description of every data system's "purpose and function." Rather, the "purpose and function" of a data system may vary depending on the team that is using it. Many data systems are used by multiple teams.

8. Last, my team has not analyzed the "types of user data" within each data system, which can also vary depending on the team using the system.

9. Based on my experience, gathering these additional data points about each of the 149 data systems we have identified as containing user data would likely require us to repeat the same process we implemented to compile this inventory, including spending more than a year revisiting each data system with the relevant departments and employees.

10. My team and others have separately begun a process of inventorying specific data assets (*e.g.*, specific data sets or logs). While we have collected metadata for some categories of assets using automated tools, we have not completed this project. Indeed, we still need to develop specific mechanisms or tools for doing so, as many of our data systems cannot be assessed using the automated tools we presently have access to.

11. I understand from my colleagues in engineering that there is no existing tool that can be readily re-purposed to identify and retrieve data related to a specific user from across all data systems. Rather, I understand it would take more than a year of work by a full team of engineers to attempt to develop a tool for that purpose.

I declare under penalty of perjury that the foregoing is true and correct, and that I executed this Declaration on December 10, 2021, in Belmont, California.

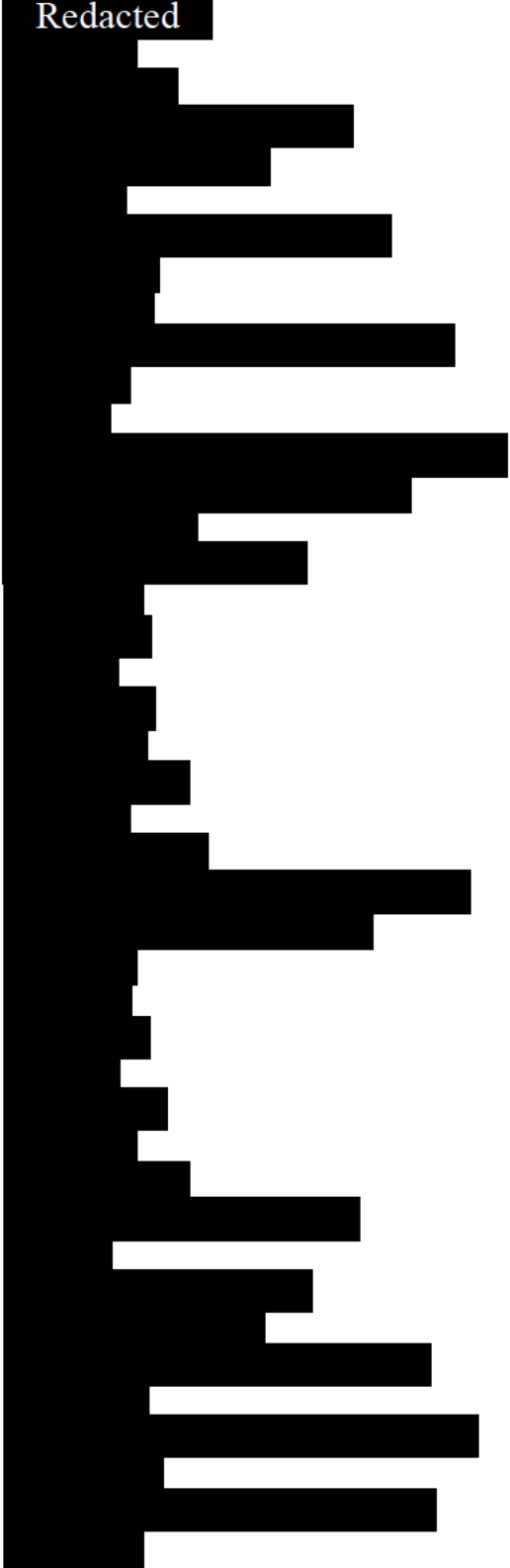
A handwritten signature in brown ink, appearing to read "David Pope", is written over a horizontal line.

David Pope

**Exhibit A to Pope Declaration**

**Data Systems That May Store or Interact With User Data**

Redacted



Redacted

[REDACTED]

Redacted

Redacted



# **EXHIBIT C**

GIBSON, DUNN & CRUTCHER LLP  
Orin Snyder (*pro hac vice*)  
osnyder@gibsondunn.com  
200 Park Avenue  
New York, NY 10166-0193  
Telephone: 212.351.4000  
Facsimile: 212.351.4035

Kristin A. Linsley (SBN 154148)  
klinsley@gibsondunn.com  
Martie Kutscher (SBN 302650)  
mkutscherclark@gibsondunn.com  
555 Mission Street, Suite 3000  
San Francisco, CA 94105-0921  
Telephone: 415.393.8200  
Facsimile: 415.393.8306

GIBSON, DUNN & CRUTCHER LLP  
Deborah Stein (SBN 224570)  
dstein@gibsondunn.com  
333 South Grand Avenue  
Los Angeles, CA 90071-3197  
Telephone: 213.229.7000  
Facsimile: 213.229.7520

Joshua S. Lipshutz (SBN 242557)  
jlipshutz@gibsondunn.com  
1050 Connecticut Avenue, N.W.  
Washington, DC 20036-5306  
Telephone: 202.955.8500  
Facsimile: 202.467.0539

*Attorneys for Defendant Facebook, Inc.*

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION**

IN RE: FACEBOOK, INC. CONSUMER  
PRIVACY USER PROFILE LITIGATION,

This document relates to:

ALL ACTIONS

CASE NO. 3:18-MD-02843-VC

**DECLARATION OF MENGGE JI IN SUPPORT OF FACEBOOK, INC.'S  
MOTION FOR RECONSIDERATION OF  
THE SPECIAL MASTER'S ORDER RE  
PLAINTIFFS' MOTION TO COMPEL  
PRODUCTION OF PLAINTIFF DATA**

Discovery Special Master Daniel Garrie, Esq.

**DECLARATION OF MENGGE JI IN SUPPORT OF FACEBOOK, INC.'S  
MOTION FOR RECONSIDERATION OF THE SPECIAL MASTER'S ORDER RE  
PLAINTIFFS' MOTION TO COMPEL PRODUCTION OF PLAINTIFF DATA**



I, Mengge Ji, declare as follows:

1. I am a Data Scientist at Meta, Inc. f/k/a Facebook, Inc. (“Facebook”). My job responsibilities include, among other things, understanding and working with Facebook’s data systems, writing queries and conducting analyses of these data, researching Facebook’s data and related technologies, and locating, analyzing, and exporting data for production in litigation and other legal matters. I submit this declaration in support of Facebook’s motion for reconsideration of the Special Master’s Order re Plaintiffs’ Motion to Compel Production of Plaintiff Data. Unless otherwise stated, I have personal knowledge of the facts set forth herein, and, if called as a witness, I could and would competently testify thereto.

2. In my role as Data Scientist at Facebook, I am familiar with the general categories of user information Facebook maintains, where that data might be stored, and how they can potentially be accessed.

**Data Contained in Facebook’s Social Graph Related to Users’ Activities on Facebook**

3. Facebook users directly provide Facebook with data or information when using the Facebook platform. For instance, a user can upload a photo directly to her Facebook profile, which Facebook retains in order to display when that profile is accessed (if the user’s privacy settings allow it). If the user “tags” a friend in that photo, Facebook also retains a record of which other Facebook user has been identified as appearing in that photo. In essence, the entire Facebook product that users interact with is a web of data points and relationships between data points that Facebook’s systems present in a user-friendly format.

4. Facebook uses the term the “Social Graph” to describe this complex web of people, places, things, actions, and connections on the Facebook platform. As Facebook users navigate through Facebook and interact with it—including, for example, by commenting on

posts made by other users, watching videos, posting photos, and sending messages—the users create new relationships and connections between themselves and the content they are able to see. The activities a user takes on the Facebook platform, including posting pictures to their profiles, tagging friends in photos, and commenting on friends’ timelines, are reflected in the Social Graph.

5. The Facebook product that users see is powered by a series of databases that work in tandem to provide Facebook users a seamless experience.<sup>1</sup> The key databases Facebook uses to support the Facebook product, which store the user content and information presently accessible via the Social Graph, are:

- **MySQL:** MySQL relational databases store much of the data that underlies the Facebook products.
- **TAO:** TAO (short for “The Associations and Objects Server”) is a service for storing, caching, and querying “associations” (*e.g.*, a like) between “objects” (*e.g.*, a user and a news-feed post). It plays an integral role in providing content to users of Facebook’s various products.
- **ZippyDB:** ZippyDB is a database used for storing logging data, derived data, and service data.

---

<sup>1</sup> Facebook’s Social Graph is not powered by a single database. Rather, Facebook is powered by an extraordinarily complex information architecture that stores information in various databases. The information in these databases is generally not human-readable and instead is intended to be processed for human consumption through Facebook’s production environment.

- **Everstore and Manifold:** Everstore and Manifold are binary large object storage solutions that store photos, videos, and other large, static files for production across Facebook's products.<sup>2</sup>

6. Between them, these databases contain trillions of data points provided to Facebook by its more than 2 billion users, many of which need to be readily accessible at all times so that users can view them as they use the Facebook product. As a result, the data infrastructure underlying the Social Graph is not only extraordinarily complex, but has also been specifically designed to serve the needs of the Facebook product.

7. One of the functions of the Facebook product is to be able to specifically identify user profiles or other sets of information associated with specific users. Put simply, when a Facebook user tries to access a friend's profile, she is essentially searching the Social Graph for certain types of data associated with that user. The resulting data is then processed through Facebook's production environment so it is presented to the user in the form of a Facebook profile, rather than a table of machine-readable data.

8. In order to support these and similar functions, the Social Graph has been engineered so that data relating to particular users can be retrieved quickly in the production environment. Through APIs, engineers and developers can query the Social Graph for data associated with a specific user ID. The results of the query are then rendered through Facebook's production environment to be human-readable.

9. I understand that Facebook's "Download Your Information" or "DYI" tool retrieves data from and allows users to download a copy of data Facebook associates with their

---

<sup>2</sup> The Everstore database is in the process of being deprecated. The data contained in Everstore is being transitioned to Manifold.

Facebook account, including data associated with their account in the Social Graph. In order to compile this information, the DYI tool runs a set of searches that have been engineered by the DYI product team in order to pull data associated with that user from the Social Graph. The resulting data is presented in a user-readable format, rather than as code or strings of data.

**Analytics Data Stored in Facebook's Data Warehouse ("Hive")**

10. Facebook also stores data sets that it uses for internal analytics, product development, and other business functions. As discussed below, these data sources are not intended to be analyzed or used at the individual user-level and are not structured to allow centralized searches for an individual user's data. Rather, searches for data related to an individual user would be an ad hoc, time-consuming, and complex process that would yield raw data that may not be human-readable.

11. Facebook's internal analytics data is primarily stored in a data warehouse called Hive, which exists separately from the Social Graph. Hive stores logs of certain activity and snapshots of entities from the Social Graph (*e.g.*, the number of unique sessions on Facebook each day) if and only if a Facebook employee chooses to write code to log that information. In addition, other types of data not relating to users, such as diagnostic or operational data relating to the performance of products and tools, may be transferred to Hive for storage purposes.

12. The Hive data warehouse is contained across a decentralized set of data centers, spread across the world. Hive contains over 20 million tables that take up more than 12 exabytes of logical storage space, all stored in a highly-compressed form.

13. In addition to the Social Graph and Hive, I am aware of several other databases that contain data related to Facebook users, including data that is anonymized. I am not aware of any additional databases where it is possible to retrieve all data associated with a particular user

without separately searching for that user within each of the data sets in the database. As a result, searching these databases for individual user data would be a time-consuming, burdensome, manual, and iterative process akin to that described below.

14. I have also consulted with other members of the Data Science team regarding this question. My colleagues are also not aware of databases containing user data other than those underlying the Social Graph that are indexed or searchable by user.

15. As detailed below, it would be extraordinarily burdensome to identify data associated with a specific user from data sources that are not indexed or searchable by user, like Hive.

### ***Searching Within Hive***

16. Unlike the Social Graph, Hive is not indexed by user and cannot be centrally searched for all data associated with a specific user ID. In other words, a Facebook data scientist cannot simply run a search for a specific user ID and have a list of all data points from all Hive tables that relate to that user ID, or even a list of data tables that contain that user ID. A separate query would need to be written to search each table for data pertaining to specific users. Tables are organized into more than 200 individual databases roughly based on product, team, or use case. Otherwise no centralized categorization exists across all of Hive. While some highly-utilized tables are documented and maintained by particular teams at Facebook, many other tables are not documented nor maintained at all.

17. The data within Hive is so voluminous that Facebook data scientists typically seek the advice of subject matter experts and employees in relevant departments before attempting to identify specific categories or types of data. For instance, if a data scientist was endeavoring to analyze data related to the performance of a certain product, the data scientist would consult that

product team to identify Hive tables they typically rely on or use to track data and that would be helpful to that analysis. It is typically not efficient or productive to endeavor to just search randomly within Hive for categories or types of data.

18. Facebook does not have a process or tool that would allow it to identify all data points associated with a specific user within Hive. Rather, doing so would be a manual process: a Facebook data scientist would need to individually assess and search millions of Hive tables for a specific user ID or set of user IDs. This would be an extremely time-consuming process: I would first need to assess each table to determine whether it contained user data, which would take approximately 3 minutes per table, or 6 million minutes if I only assessed 10% of the tables in Hive. I would then need to search each table for data related to the specific users at issue. Depending on the size of the table, this search could take as little as a few minutes, or as much as 1-2 weeks, to run. Even if this search took only 1 hour to run per table, searching even 2.5% of the Hive tables would take 500,000 hours, or more than 200 years of full-time work by a Facebook data scientist, only to run the searches, not analyze or export any of the data.

19. By way of analogy, if Facebook were a library and each of the millions of Hive tables were a book within that library, Facebook would not be able to search through all of the books at once to identify how many contained a particular phrase. Rather, Facebook would have to open and read each book individually to determine whether it contained that phrase. The search for data relating to a specific user within each of the millions of Hive tables would be similarly manual and onerous. This would be a monumental undertaking.

20. Facebook is also limited in its ability to identify data associated with particular users, as within 90 days of logging data associated with individual users in Hive is generally disassociated from the user's identifiable information, anonymized, or deleted (depending on the

nature of the data and any business reasons for retaining it). Even in instances where a data scientist can trace disassociated data back to specific users, this additional step must still be run manually across millions of data sets.

21. This search is also time-consuming due to the fact that many Hive tables are stored in “partitions” or segments, the size of which can vary depending on the size of a data table. Many tables are partitioned by day, such that each day of data is its own segment of the larger table. Because each segment is partitioned off from the others, Facebook cannot quickly or efficiently run searches, for instance, for particular user IDs, across multiple days of data in a partitioned table at once. Instead, Facebook’s system is optimized for running searches within an individual partition.

22. Many fields in Hive tables do not contain human-readable information and instead contain numerical values. As a result, if the information needs to be used or analyzed outside of Facebook’s environment, the numerical values would need to be cross-referenced with other tables or data legends to be understood. This is an extra step of analysis that would need to be conducted if the results of any searches for specific user IDs would need to be produced to the Plaintiffs. This would add between a few minutes and a few days of work to each search that yielded results.

23. Even after cross-referencing the Hive results with other tables or data legends, the output of searching the Hive table would be primarily raw data and difficult to interpret. Within Facebook, this data is generally analyzed by specially trained Facebook data scientists. Nor could the results of searching individual tables be easily consolidated. Because the search is conducted separately for each table, the results would be output in at least as many files as the number of tables searched.

*Exporting Data from Hive*

24. Even if Facebook were to spend several years searching for all data points associated with specific user IDs, this data could not be readily exported or produced to Plaintiffs. The Hive is designed to ingest and process data for internal analysis, not to export data into files that can be transferred outside of Facebook.

25. For Facebook to export and produce data from Hive tables, the data must be decompressed and written to .csv files (which can be opened with Excel). Particularly large datasets would also need to be broken into small pieces prior to being decompressed and written to .csv files. This would be a process conducted by Facebook data scientists on the legal team. This manual, iterative process introduces the possibility of human error, which necessitates real-time quality checks of the data downloads that take additional time and create the potential for delay.

26. First, a Facebook data scientist must write entirely new code to query the data and write it into .csv files.

27. Second, the code is run to export that data to a server, which involves querying the data warehouse and then writing the results of that query as .csv files in a decompressed flat-file format. The speed at which Facebook can export the data from the warehouse to the server varies depending on usage of its systems in the ordinary course of business, but based on testing and prior exports, Facebook's current best estimate is that it takes about 40 minutes to write a 50 GB .csv file.

28. Third, once the server is full, a data scientist must empty the server by compressing the .csv files and transferring them into a local encrypted hard drive or a secure transfer site. The process must be repeated over and over until the export is complete.



29. Preparing any productions also requires substantial additional resources from senior members of the eDiscovery data scientist team to prepare for and manage the coding and export project. Moreover, after the initial exported files have been created, it would take additional time for a review team to validate the queries and data and perform additional quality control, or longer if there are issues that require remediation.

30. In total, Facebook currently estimates that, even without any other responsibilities or deadlines, it would be a substantial, multi-year project to attempt to search through millions of Hive tables for data associated with specific users and then export that data for production through the process described above. The Facebook eDiscovery team, which is responsible for assisting Facebook in-house and outside counsel in active litigations and other legal matters, in addition to building and maintaining internal infrastructure crucial to the management and preservation of data on legal hold, does not have the time and resources required to search for, access, analyze, and export the data in millions of Hive tables in the manner described above.

31. Facebook cannot materially shorten this timeline by hiring new employees because it would take time and resources to interview, select, and onboard new employees, and any new hires would have to be sufficiently trained regarding Facebook's systems, policies, and procedures, which is a mandatory process that takes three weeks. For the same reason, Facebook cannot simply engage third party consultants or temporary employees to handle this data export. Nor would adding more servers—which would require diverting them from their use in the ordinary course of business—necessarily reduce the estimated timeline in a linear fashion, as the export still requires downloading the data from the server, and the manual process of writing the code and monitoring the export. All of these options—hiring new employees, hiring contractors, and adding servers—would also be extremely costly.

I declare under penalty of perjury that the foregoing is true and correct, and that I executed this Declaration on December 10, 2021, in Sausalito, California.



---

Mengge Ji

# **FACEBOOK OPPOSITION EXHIBIT C**

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION**

IN RE: FACEBOOK, INC. CONSUMER  
PRIVACY USER PROFILE LITIGATION,

This document relates to:

ALL ACTIONS

CASE NO. 3:18-MD-02843-VC-JS

**DECLARATION OF  
BEN MITCHELL IN SUPPORT OF  
FACEBOOK'S OPPOSITION TO  
PLAINTIFFS' MOTION TO COMPEL  
NAMED PLAINTIFF DATA**

I, Ben Mitchell, declare:

1. I am Director of Product Management at Defendant Facebook, Inc. (hereinafter “Facebook”). I make this declaration on my own knowledge, and I would testify to the matters stated herein under oath if called upon to do so.

2. In my role as Director of Product Management, my responsibilities include providing support for Facebook’s “Download Your Information” or “DYI” tool. Through my role, I am familiar with the DYI tool, the data that it includes, and where that data is stored.

3. Facebook uses the term the “Social Graph” to describe the complex web of people, places, things, actions, and connections on the Facebook platform. The Facebook product that users see is powered by a series of databases that work in tandem to provide Facebook users a seamless experience. As Facebook users navigate through Facebook and interact with it—including, for example, by commenting on posts made by other users, watching videos, posting photos, and sending messages—the users create new relationships and connections between themselves and the content they are able to see.<sup>1</sup> This web of people, places, things, actions, and connections is referred to as the “Social Graph.”

4. The DYI tool allows a user to download a copy of data Facebook associates with their Facebook account, including data associated with their account in the Social Graph.

5. The DYI file for each individual user represents the most complete and best compilation of data Facebook maintains associated with that user, and the best available

---

<sup>1</sup> Facebook’s Social Graph is not powered by a single database. Rather, Facebook is powered by an extraordinarily complex information architecture that stores information in various databases. The information in these databases is generally not human-readable and instead is intended to be processed for human consumption through Facebook’s production environment. I am generally aware that several databases collectively store the information that underlies the Social Graph and the names of these databases, but I am not knowledgeable about the technical details and functions of each underlying database.

compilation of the data about that user in the Social Graph, in a human-readable and producible form.

6. Exhibit B to the Declaration of Martie Kutscher lists categories of data contained in a user's DYI file, unless data in a given category does not exist for the user or the user deleted it. The DYI file does not include data such as (i) data a user has deleted from their own profiles (e.g., photos that have been removed), (ii) any other data Facebook does not maintain; (iii) data associated only with a different user's account, or (iv) Facebook's trade secrets.

I declare under penalty of perjury that the foregoing is true and correct. Executed on October 28, 2021 at Felton, California.

*Ben Mitchell*

---

Ben Mitchell

# **FACEBOOK OPPOSITION EXHIBIT E**

**UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION**

IN RE: FACEBOOK, INC. CONSUMER  
PRIVACY USER PROFILE LITIGATION,

This document relates to:

ALL ACTIONS

CASE NO. 3:18-MD-02843-VC-JS

**DECLARATION OF  
KARANDEEP ANAND IN SUPPORT OF  
FACEBOOK'S OPPOSITION TO  
PLAINTIFFS' MOTION TO COMPEL  
NAMED PLAINTIFF DATA**



I, Karandeep Anand, declare:

1. I am Vice President, Business Products at Defendant Facebook, Inc. (hereinafter “Facebook”). I make this declaration on my own knowledge, and I would testify to the matters stated herein under oath if called upon to do so.

2. In my role as Vice President, Business Products, I am familiar with how Facebook shares or makes data available to third parties, including application developers and partners. I am also generally familiar with how Facebook’s platform operates, how data about particular users can be accessed, and how that data is shared with third parties.

3. Facebook makes individualized data about Facebook users available to third parties—including app developers and partners—through application programming interfaces (“APIs”). These APIs pull data exclusively from Facebook’s Social Graph.

4. Facebook uses the term the “Social Graph” to describe the complex web of peoples, places, things, actions, and connections on the Facebook platform. The Facebook product that users see is powered by a series of databases that work in tandem to provide Facebook users a seamless experience. As Facebook users navigate through Facebook and interact with it—including, for example, by commenting on posts made by other users, watching videos, posting photos, and sending messages—the users create new relationships and connections between themselves and the content they are able to see.<sup>1</sup> This web of people, places, things, actions, and connections is referred to as the “Social Graph.” The activities a user takes on the Facebook Platform, including posting pictures to their profiles, liking photos, and

---

<sup>1</sup> Facebook’s Social Graph is not powered by a single database. Rather, Facebook is powered by an extraordinarily complex information architecture that stores information in various databases. The information in these databases is generally not human-readable and instead is intended to be processed for human consumption through Facebook’s production environment. I am generally aware that several databases collectively store the information that underlies the Social Graph and the names of these databases, but I am not knowledgeable about the technical details and functions of each underlying database.

commenting on friends' timelines, and certain activities a user takes off of the Facebook Platform, are reflected in the Social Graph.

5. APIs are a standard industry programming tool and they allow applications to access data and features of other applications, services, or operating systems. APIs can provide access to a defined set of User Data (e.g., a user's name or Facebook ID) or other information the developer is authorized to access, (e.g., photos a user has shared with the developer).

6. APIs Facebook has made available to third parties—including app developers and partners—that allow access to user-identifiable information query the Social Graph only. This was also true during the period from 2007 to present and is true of all of the APIs identified in Facebook's response to Plaintiffs' Fourth Set of Interrogatories. Users' privacy settings and the permission they specifically grant the applications they access or download control what non-public data third parties are able to access about them, as described in Facebook's Data Policy.

7. Facebook also maintains a data warehouse called Hive, which is separate from the Social Graph. Hive stores more than 12 million tables, including logs of certain activity from the Social Graph (e.g., the number of unique sessions on Facebook each day) if a Facebook employee writes code to log that information. Facebook does not provide any APIs that allow third parties to retrieve data from Hive and third parties are not able to access Hive directly. Hive is not indexed by user, and there is no way to search Hive by user—each individual Hive table must be searched individually.

I declare under penalty of perjury that the foregoing is true and correct. Executed on October 28, 2021 at Seattle, Washington.

*Karandeep Anand*

---

Karandeep Anand